# From Clinical Data to Discovery: integration of Beacon v2 with OMOP-CDM

Sergi Aguiló[1], Alberto Labarga[1], Miguel Ángel Mayer[2], Juan Manual Ramírez-Anguita[3], Oriol López-Doriga Sagales[4], Aurora Moreno-Racero[4], Liina Nagirnaja[4], Dmitry Repchevski[1], Salvador Capella-Gutierrez[1] and Jordi Rambla[4]

[1]*Barcelona Supercomputing Center (BSC), Barcelona, España*
[2]*Hospital del Mar Research Institute (IMIM), Barcelona, España*
[3]*Universitat Pompeu Fabra (UPF), Barcelona, España*
[4]*Centre for Genomic Regulation (CRG), Barcelona, España*

**Abstract**

The growing volume of clinical patient data offers unique opportunities for understanding diseases using Real World Data (RWD). However, standardized storage, discovery, and access methods that preserve patient privacy are crucial. OHDSI's Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) has become a widely used standard for organizing and analyzing healthcare data in an interoperable format, enabling large-scale observational research. Beacon v2, an open-source data discovery protocol, facilitates secure data discovery across institutions by querying OMOP CDM databases in a federated manner. Developed under the Global Alliance for Genomics and Health (GA4GH), Beacon v2 supports interoperability while protecting patient data. Beacon v2 is built on two core concepts: the Framework, which defines query mechanisms, and the Model, which structures the data. This decoupling allows flexibility in integrating various data sources like OMOP CDM or HL7 FHIR. Beacon v2 leverages ontologies to perform queries, enabling harmonization at the API level without altering underlying databases. A new Beacon v2 Production Implementation (B2PI) simplifies its deployment, especially for clinical datasets, and a specialized Beacon4OMOP extension aligns OMOP CDM data with Beacon models. Approved as a GA4GH standard in 2022, Beacon v2 enables parallel querying of unified networks of biomedical centers, returning aggregated responses while maintaining privacy. This federated approach enhances collaboration, ensuring interoperable and reusable data discovery. By integrating OMOP CDM databases, Beacon v2 supports better decision- making in healthcare and improved patient outcomes through secure and efficient data utilization.

**Keywords**

Beacon v2, OMOP-CDM, federated data discovery, data sharing

## 1. Introduction

The increasing volume of patient data generated in clinical context represents an unique opportunity to further understand diseases using Real World Data (RWD). However, before using this data for research purposes, there is an urgent need for standardized ways to store, discover, and access it while preserving patient privacy. OHDSI (Observational Health Data Sciences and Informatics) has become a key stakeholder in the use and reuse of health data at large scale. Indeed, one of the widely used approaches for organizing patient data in an interoperable format is the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM1) which is an open community data standard led by OHDSI. OMOP CDM enables observational data to be converted into a standardized database structure for organizing and analysing healthcare data from different sources. OMOP CDM facilitates large-scale observational research and easy integration with tools and formats. Beacon v22 is an open- source data discovery protocol that can query OMOP CDM databases and allow the presentation of the underlying data in a federated manner, thus enabling secure and efficient data discovery across institutions. Beacon v2 is an accepted standard of the Global Alliance for Genomics and Health (GA4GH3), which is an organization that aims to facilitate secure use of genomic and other related health data by supporting

*SWAT4HCLS 2025: 16th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences 2025*
0000-0003-0830-5733 (S. Aguiló); 0000-0001-6781-893X (A. Labarga); 0009-0003-7476-7319 (O. L. Sagales); 0009-0003-7476-7319 (A. Moreno-Racero); 0000-0002-0309-604X (S. Capella-Gutierrez); 0000-0001-9091-257X (J. Rambla)

the development of technical standards and tools. Moreover, Beacon v2 supports security protocols that enable clinical data owners to navigate the delicate balance between data interoperability, accessibility and patient data protection.

## 2. Methods

Beacon v2 is built around two core concepts: the Framework (i.e. the syntax) that describes the mechanisms of queries and parsing the responses, and the Model (i.e. the semantics) that provides the structure for the data model. This decoupling of the Framework from the Model allows flexibility in the data type and integration with various data sources, such as OMOP CDM or HL7 FHIR. While being a discovery standard by itself, Beacon v2 is built upon other GA4GH standards. The phenotypic data model is largely built on the Phenopackets4 ontology framework, whereas the representation of genomic variation is tightly linked to the VRS1.35 format.

Beacon v2 largely leverages the use of ontologies to perform the queries, and the specification itself includes suggestions. Every beacon instance declares which ontology terms are accepted in the queries. The use of ontology terms is independent of the actual terminology used in the database.Relying on ontology terms allows beacon instances to include and accept synonym terms, e.g. a beacon could accept both the terms "PAT0:0000383 - female" or "UBERON:0003100 – female organism" and translate it to the same OMOP concept. This is the basis for harmonizing at source but without requiring the transformation of database entries, just in doing and applying the mapping at the Beacon API interface level.

The implementation of Beacon v2 API requires a higher level of technical expertise, while a recently released (Oct 2024) 'out-of-the-box' implementation (Beacon v2 Production Implementation, B2PI6) offers an easier approach for 'beaconizing' biomedical datasets, particularly in clinical entities. Moreover, a data model extension, Beacon4OMOP7 has been developed for B2PI application that integrates with any OMOP-CDM database. Expert mapping between OMOP CDM and Beacon models ensures alignment of items such as diagnoses, treatments, lab results, exposures, and biological sample data.

## 3. Results

The largest utility of data discoverability is achieved with a unified network of biomedical centers contributing to the benefit of research and patient healthcare. Following its approval as a GA4GH standard in 2022, Beacon v2 discovery protocol has been widely accepted and integrated into unified networks of biomedical centres by consolidating individual beacon instances under a central Beacon v2 API. This central node allows parallel querying of all Beacons and returns an aggregated response via a unified user interface. This facilitates discoverability of the clinical data stored in OMOP CDM, while ensuring patient privacy.

## 4. Conclusion

Beacon v2 represents a secure federated data discovery approach that fosters interinstitutional collaboration and research. By incorporating OMOP CDM databases into unified Beacon networks, data is discovered in an interoperable and reusable way while giving data owners full control over the access and security measures. With the improved data infrastructure, better decision-making in healthcare institutions can be achieved, thus ultimately returning the benefits to the patients themselves

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

## 5. References


1. https://www.ohdsi.org/data-standardization/
2. https://docs.genomebeacons.org/
3. https://www.ga4gh.org/
4. http://phenopackets.org/
5. https://www.ga4gh.org/product/variation-representation/
6. https://github.com/EGA-archive/beacon2-pi-api
7. https://gitlab.bsc.es/impact-data/impd-beacon_omopcdm